SOUNDMARKS IN SPOKEN ROUTE GUIDANCE

Anssi Kainulainen, Markku Turunen, Jaakko Hakulinen, Aleksi Melto

University of Tampere, Tampere Unit for Computer-Human Interaction
Speech-based and Pervasive Interaction Group
Tampere, Finland
{firstname.surname}@cs.uta.fi

ABSTRACT

Route guidance is an emerging mobile computing application domain. Soundscapes or acoustic environments are a perceptually important part of people's location awareness and navigation. In this paper, we present how non-speech audio can be used to complement speech-based and graphical route information in mobile public transport guidance. We present TravelMan, a mobile multimodal pedestrian and public transport route guidance application. Based on TravelMan, we also present a soundmark-based route description design. Auditory icons describe methods of transport and identify spatial points of interest. They support users as a less intrusive, awareness supporting information source. Initial test results indicate that combining speech and non-speech sounds are not a trivial task, and that there is need for further development.

[Keywords: Soundmarks, Speech User Interfaces, Public Transport, Pedestrian Route Guidance, Soundscapes]

1. INTRODUCTION

As mobile devices, such as mobile phones, have become more powerful, their applications have become more elaborate and varying. The mobile context brings new possible application domains, for example navigation assistants. At the same time, the mobile use context brings limitations to interaction; users' hands and eyes are busy, they have to divide their attention between multiple tasks and applications, and the devices used for communication are small and have minimal keyboards, pointing devices, or displays. Speech and non-speech audio provide solutions for many of these challenges. Even the smallest devices can have audio input and output capabilities, and audio can be used in the background to keep users aware of dynamic information, e.g., to provide contextual information and alert users when something important happens. Route guidance is an example of a fairly modern mobile computing application. Positioning techniques, such as GPS (Global Positioning System), have become widely available to consumers. In the area of car navigation, GPS navigators with speech user interfaces are widely spread applications, and there are similar products targeted for other usage, such as sea and pedestrian navigation. Public transportation guidance applications can be seen as the next developmental step. Such services, combined with pedestrian guidance, could be of great importance to many. In particular, when carefully designed, they can be very helpful for special user groups such as visually impaired users.

A perceptually important part of people's location awareness and spatial orientation is the soundscape or acoustic environment they reside in. Soundscapes consist of three basic elements: keynote sounds, signals and soundmarks [1]. Keynote sounds provide an ongoing background identity of a soundscape, e.g., traffic in cities, but may not always be even consciously heard. Signals are finite foreground sounds, which grab the attention of the listener and often prompt some action, e.g., car horns. Soundmarks are unique to an area, just as visual landmarks, and thus can be reliable anchors for positioning oneself in the area.

In this paper we present how non-speech audio can be used to complement speech-based and graphical route information in a mobile public transport guidance application. In addition to guiding users with speech, auditory icons can be used to describe route information, such as used methods of transport and temporal information. Soundmarks can also be used to identify spatial points of interest, and provide landscape and landmark context for the navigation. The auditory icons can complement visual and speech-based guidance and support users as a less intrusive, awareness supporting information source.

In the rest of the paper we present how soundmarks can be used to support awareness in route guidance. We introduce an initial design of such an application based on the existing spoken and multimodal route guidance application. Finally, we discuss design issues of such applications.

2. SOUNDMARKS TO SUPPORT AWARENESS IN ROUTE GUIDANCE

Route guidance applications can benefit from non-speech auditory awareness information. For example, there can be a lot of contextual information in route guidance, such as non-landmark places, that provide peripheral awareness information, but are not mandatory for successful interaction. However, this "additional" information may help users to gain and maintain awareness of the route.

There is a lot of work done in the area of route guidance applications. Most importantly, commercial car navigation systems employ successfully three-dimensional graphics and spoken instructions (usually recorded human voices). There have been numerous research prototypes that study pedestrian guidance in various settings. Many of them focus on maps and graphical presentations. In general, landmarks have been identified as the most useful information for pedestrians with

normal sight. For example, photographs have been used to depict identifiable landmarks along the route [2], and speech-only instructions have been used mainly in services targeted for visually impaired users. Music and spatial audio have been also used to convey guidance information [4].

Soundmarks can be used to complement spoken and graphical guidance in a multimodal public transportation route guidance application. We present a design of sound presentations to help users to understand the structure of the route, different modes of transportation, and key locations and events, such as bus stops, crossroads and landmarks which will help identify them. These descriptions are designed to help the users in navigation, to select correct routes and vehicles, and when and where to get in and out of the vehicles. These soundmarks aim to be a subtler manner of supporting awareness in route guidance tasks compared to explicit spoken interaction and long descriptive phrases. The auditory awareness support is especially valuable for people with disabilities, e.g., visually impaired. Next, we present our route guidance application, which utilizes soundmarks.

2.1. Mobile Multimodal Route Guidance Application

TravelMan is a multimodal mobile application that provides transport information services in Finland. The application is based on the research on spoken and multimodal transport information systems developed in a Finnish research project (http://www.cs.uta.fi/hci/spi/TravelMan/). The main functionality of the application is to provide route guidance information for public transport, such as subway, tram, and bus traffic in Finnish cities. In addition, information for long-distance traffic is included. Here we focus on Helsinki metropolitan area local traffic information.

Figure 1. *TravelMan application.*

Figure 1. illustrates the application. All information is both spoken and displayed on the screen. They are, however, not the same on linguistic level. Instead, output content is optimized for each modality. In the case of spoken outputs this means, for example, that complete sentences are used, and to maximize intelligibility and pleasantness, word choices are in some cases different. Similarly, speech and keypad inputs can be given to the system. Telephone keypad is used for navigation in menu structures. The menu structures employ a two-dimensional reel metaphor, where menu items and sub-routes are placed on top of each other sequentially. Reel navigation is supported by haptic and auditory feedback, which imitate flipping through a deck of cards or a Rolodex. Text input can be used for entering names and addresses of departure and destination places.

Because of the speech outputs and the reel user interface, the application can be used without seeing the screen. After finding out a suitable route, the user can put the mobile phone into his/her pocket, listen how the journey progresses, get tactile feedback, and give keypad inputs. This makes the system accessible to visually impaired.

The application supports GPS devices, so contextual real-time guidance can be provided. For example, when the user has turned a hands-free on, TravelMan will provide information about the progress of the journey, where to step on and off a bus, and where to go next. Figure 2. has an example, which can also be listened:

http://www.cs.uta.fi/hci/spi/TravelMan/audio/fountain.ogg

Our design works as a dynamic guidance system with GPS support, or as a referential route description guide without such support. We believe route guidance is possible even with relying only on non-position-aware route descriptions and soundmarks, though this remains to be proven in our next implementation. Adding a hands-free headset makes the sound quality good enough for using high fidelity soundmarks, and relieves users' hands for easier operation, e.g., when keeping one hand in a pocket to control the phone. In addition, the use of a headphone is an important aspect for the social acceptability of the application.

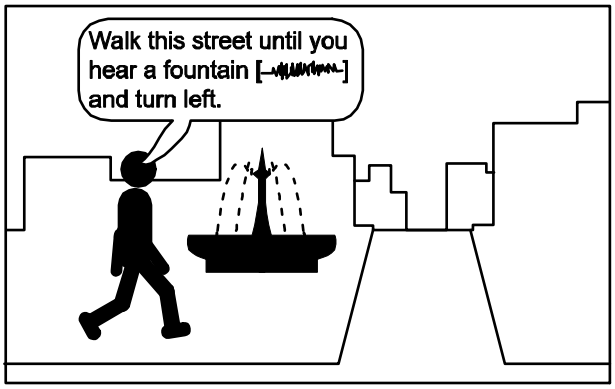

Figure 2: *TravelMan guidance scenario.*

3. INITIAL AUDITORY PRESENTATIONS

Based on the route guidance application presented, we designed an initial version of audio route descriptions. The application's route descriptions are based on a national public transportation guidance system (http://www.journey.fi/). Each route description offered by our application consists of a varying number of vehicles, and varying lengths of sub-routes, as depicted in Figure 3. The lengths are given as temporal distance (duration) between locations, instead of spatial distance. In this phase we focus on auditory icons that provide basic awareness information with sounds of vehicles. In the next phase we add more

awareness information, such as soundmarks representing important places and objects (e.g., landmarks) in the routes.

We created three different styles for auditory route descriptions: spoken, non-speech and a combination of both. Sound examples of these are also available:

http://www.cs.uta.fi/hci/spi/TravelMan/audio/spoken.ogg
http://www.cs.uta.fi/hci/spi/TravelMan/audio/nonspeech.ogg
http://www.cs.uta.fi/hci/spi/TravelMan/audio/combined.ogg

Spoken presentations are based on the route guidance application, as illustrated in Figure 1, and designed to take advantage of a high-quality phoneme based synthesizer 'Concept-to-speech' developed at the University of Helsinki [3]. The spoken descriptions tell the mode of transportation, target for each separate sub-route and its duration, e.g., "By walking to Hämeentie-road number thirty-one, two minutes. (pause) By metro to East Central, ten minutes." Since speech is a quite slow output medium, the duration of spoken outputs vary considerably, depending on the addresses and complexity of the routes.

Figure 3: *An example of a route with four sub-routes: 7 min walking, 9 min subway, 5 min tram, and 5 min walking.*

Non-speech presentations are built from auditory icons of four modes of public transportation in the Helsinki metropolitan area: the sounds of walking, metro trains, trams and buses. We recorded the vehicles on location around the city, from within the vehicles (e.g., as people hear them when traveling). The recordings were used to pick representative sounds elements, from which the auditory icons were built of. Mono sounds are used, since we do not see spatial imaging viable due to implementation platform and usage context. Our design for route guidance is suitable for large amounts of users, since it does not require special equipment in addition to a mobile phone and an optional hands-free set. Spatial audio might give more precise guidance, but requires stereo or binaural headphones, which are a potential hazard while moving in city traffic.

For each mode of transportation, an accelerating, constant speed and decelerating sound was chosen. The acceleration represents a starting vehicle, and deceleration represents a stopping vehicle. The tempo of each sound is doubled to make them more iconic and discernable from actual traffic sounds. Pitch is not conserved, so it changes with the tempo. After the pitch and tempo change, the acceleration and deceleration sounds were edited to last half a second long each. The constant speed sound samples are longer, and loopable.

These sound elements are combined to construct route descriptions. A single route description consists of the sound of a vehicle accelerating, running at constant speed for a set time, then decelerating, after which the same treatment is done with the next mode of transportation. The auditory icons represent the modes of transportation, and the duration and order they were played in represent the structure and duration of the route. Since the acceleration and deceleration are fixed to half a second each, the duration of the constant speed part represents the duration of a journey. For each minute of travel in real time, a second passes in the sound presentations. The shortest possible sub-route, one minute, thus consists only of acceleration and deceleration. Longer sub-routes keep looping the constant speed for the required additional time.

The third presentation style combines the other two. For each sub-journey, synthesized speech tells the possible vehicle line number and final stop of that sub-route, after which auditory icons are played as in the second presentation style. Auditory icons represent the duration of the sub-route, and the vehicle type used. This information is not told by speech, as it is in the first presentation style.

4. DESIGN CONSIDERATIONS

We conducted an initial test comparing the three auditory route descriptions with 57 participants. Participants were divided in six groups for counterbalancing. Each auditory description design was presented as three recognition tasks of increasing difficulty and preceded by one practice task. Each task consisted of one audio presentation followed by a multiple-choice task from four graphical presentations. Tasks were followed by an opinion questionnaire.

Speech only presentations had a mean recognition rate of 56% over all tasks, auditory icon presentations 45%, and combined presentations 34%.

Unipolar opinions gave a mean 72% out of maximum approving score for the speech-only presentation, mean 49% out of maximum approval for the combined presentation, and mean 35% out of maximum approval for the auditory icons only presentation. Bipolar opinions gave a mean deviation of 2% from neutral judgment for combined presentation, 4% mean deviation for speech only presentation, and 25% for auditory icons only.

The test showed that auditory icons weakened the overall opinion of the presentations when used together with speech, but they were still liked more than non-speech only. On the other hand, the combined presentations were the most difficult to recognize. This gives rise to the question whether combining different kinds of sound is especially difficult. Recorded "natural" sounds together with synthesized speech, even with a relatively high quality synthesis, might not fit in the same sound ecology easily. There have been similar results between recorded and synthesized speech [6], and even between different types of speech syntheses, which hint that moving from one type of sound to another is difficult. Speech and non-speech sounds could be seen as different modalities, and combining them is not trivial.

4.1. Route Descriptions

There are acoustic differences between the speech samples and the auditory icons used in our presentations. Speech samples are synthesized speech, although of relatively high quality. Auditory icons contain all kinds of unnecessary components, like echo and background noise, which makes them psychoacoustically different from the crisp and clear synthesized speech. Speech and non-speech elements are played sequentially, so the abrupt

changes in the overall soundscape might be disruptive. Layering elements on top of each other, or adding a common keynote sound could yield better results, even if speech clarity would not stay the same. It would be interesting to compare sequential and layered information presentation styles later on.

Auditory icons were recorded, but their treatments make them less natural. Because the tempo and frequency are changed, the sounds might be more discernable from actual traffic sounds of the supposed use context, but the frequency ranges are only shifted, not completely changed.

To see whether this has much effect to recognition, we analyzed the spectral qualities of two vehicles we supposed to be hardest to separate, i.e., the tram and the metro. At a constant speed, both have two main sound sources: the engines, and the wheels against the tracks. Naturally, since they run at different maximum speeds, the spectral components of engines and wheel sounds spread to higher frequencies on the metro. The engines of both vehicles produce a strong base frequency, and the wheels produce higher frequency components at even intervals. The main difference (beside the metro having generally higher frequencies) is that the spread of each frequency component on a tram is wider, which makes it somewhat softer while the metro sounds quite sharp. This quick test confirms our presumption that more finite sound events, i.e., acceleration and deceleration, opening and closing doors, assorted signals etc., are more important for the recognition of a vehicle sound. And while the frequency differences of the two vehicles are clear, altering those frequencies through tempo change might make them even less familiar, although the relative frequency component differences stays the same.

Since our vehicle icons have only two half-second samples of acceleration and deceleration, and most of the vehicle sound is steady speed sound, this could have been the worst choice in their design. Although sonifying actual travel time in this manner initially seemed logical, including more finite sound elements could improve recognition results. A noise suppressing treatment on the vehicle sounds may make the vehicles sound less like traffic noise, and make them more abstract and iconic at the same time. Although this could lower the initial recognition result, the learning curve should not be overly steep, since people experience similar noise abatement with active noise suppression stereo earphones, or within well-insulated vehicles. Another approach would be to analyze the vehicles' sounds further, and construct simple additive syntheses to imitate the vehicles. More developed methods, e.g., short-time Fourier transform –based algorithms and tracking phase vocoders can be used to control the synthesis parameters.

As a soundscape, our design's keynote sound was the sound of traffic present in the vehicle icons themselves. Their role in the initial design was not consistently implemented, especially in relation to speech elements. The beginnings and endings of each vehicle could be seen as sound signals, and they should be emphasized more, in order to carry the messages "vehicle identity", "enter vehicle" and "step off vehicle" better. In future designs, the balance between the keynote sound, sound signals, and the soundmarks to be used in guidance need to be better considered. The role of speech is challenging, since while fully articulated sentences are clear and easy to understand, they might mask the background soundscape and soundmarks too much. Shorter words as parts of the presentation might bring more balance to the presentations, but their linguistic incompleteness might also prove to be problematic. The vocabulary needs to be carefully considered, since small differences, e.g., saying "turn right, to Kalevantie-road" instead of "turn right", might give only marginal additional information while possibly being even detrimentally excessive considering the entire presentation as a whole. Static presentations could be served better with more minimalist choices, while interactive and dynamic presentations could benefit from access to more detailed information.

4.2. Soundmarks in Guidance

Soundmarks are by definition recognizable and unique to a place. Finding and gathering suitable soundmarks for guidance purposes has some challenges. Identifying unique sounds takes time and making them usable in an audio presentation takes some technical skill. Finding and editing suitable sounds creates a burden for soundmark content producers if the application is to be used on a larger scale, so a communal effort may be an answer. Sometimes, a good soundmark is not physically or visually unique, and a location can be identified through the entire soundscape instead of any individual soundmarks. For example, there might be only one place in a city where, e.g., the sounds of seagulls, railway crossing bells, and cars waiting to get across are audible together.

Some soundmarks occur only at certain times of a day, or their acoustic horizon fluctuates due to surrounding circumstances. Because of this, we believe multiple and alternative soundmarks should be made available for better coverage and reliability. Soundmark-based guidance requires a chain of locations, where soundmarks, other guidance and earlier history support each other in the ongoing rechecking of the supposed location.

For general route guidance purposes, soundmarks should be objectively unique. Often soundmarks have a communal aspect, uniqueness as defined by people who live in the area and their long-term experiences and memories. While people familiar with the area will know the locations and sounds of well-known soundmarks, the main target group of a route guidance application may be completely unfamiliar with the place.

On some occasions, an abstract spoken description of a soundmark [5] may prove more reliable than an actual recording, if the soundmark changes a lot. On the other hand, linguistic constraints may even favour using a symbolic soundmark, even if it would not be an actual recording from the location.

5. DISCUSSION AND FUTURE WORK

We presented soundmarks as a method of supporting awareness in route guidance. Based on our earlier research and prototypes, we designed and compared spoken, non-speech audio, and combined route description designs. These designs acted as a starting point for our ongoing research on an auditory route guidance application, which focuses on using soundmarks, soundscapes and speech in public transport and pedestrian route guidance.

We conducted an initial comparison test for the route description designs, but the results were not flattering and there are some questions of their reliability. They mostly show a strong need for further development of the presentations, so we did not report a more detailed analysis.

Designing auditory icons for route guidance is not trivial, but possible. The balance of sound elements in a route guidance soundscape (keynote sounds, signals and soundmarks) needs to be carefully considered, especially when including speech output. Focused and peripheral interaction need to support each other, since using especially dynamic route guidance has phases of activity and inactivity, where focus of the user's attention changes. We see that complementary spoken and non-speech auditory output can circumvent each other's linguistic or acoustic limitations, or even create a new kind of dialogue between the user and the system.

Pedestrian and public transport is a challenging use context for audio applications. While private cars are socially private places and have more freedom in interaction choices, public transport vehicles are public places with different rules for acceptable behavior. The privacy of walking on the street depends much on location and time of day. In some places the user can be the only person within hundreds of meters, while in a bus, at least the driver is always present. People are used to hearing music while sitting on buses and other vehicles, but synthesized speech is still very uncommon in general, and even less so in public places. Engaging in dialogues not resembling natural human-human dialogues might prove socially unacceptable for most people because of the attention it would attract, and the possibly private information being discussed. People are somewhat used to hearing auditory icons, e.g., in the form of mobile phone ring tones, so we believe their acceptability is somewhere between music and synthesized speech. The overall situation with privacy of information dealing with one's movements is an ambiguous one, since many people seem quite comfortable talking their personal lives to other people through phones, even though there are many unfamiliar people within earshot. Privacy problems of audio output can be dealt with earphones.

In addition to social acceptability, noisy traffic environments could prove difficult for the clarity of audio applications. Relying only on auditory icons is problematic, since within vehicles, one cannot usually hear the outside world well enough to discern soundmarks of the environment. This leads us to believe soundmarks would be most effective when walking or bicycling. Vehicles with poor sound insulation might mask the audio output of an application. Hands-free devices and other earphones would lessen this problem.

6. REFERENCES

[1] R.M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, USA, 1977 / 1994.

[2] L. Chittaro, and S. Burigat, "Augmenting audio messages with visual directions in mobile guides: an evaluation of three approaches," in *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services (Mobile HCI 05)*, Salzburg, Austria, September 19-21, 2005.

[3] M. Vainio, A. Suni, and P. Sirjola, "Developing a Finnish concept-to-speech system," in M. Langemets and P. Penjam, editors, *Proceedings of the Second Baltic Conference on HUMAN LANGUAGE TECHNOLOGIES*, Tallinn, April 4--5 2005, pp. 201-206.

[4] M. Jones, G. Bradley, S. Jones, and G. Holmes, "Navigation-by-Music for Pedestrians: an Initial Prototype and Evaluation," in *Proceedings of the International Symposium on Intelligent Environments: Improving the quality of life in a changing world*, Homerton College, Cambridge, United Kingdom, 5-7 April, 2006.

[5] J. Baus, R. Wasinger, I. Aslan, A. Krüger, A. Maier, and T. Schwartz, "Auditory Perceptible Landmarks in Mobile Navigation," in *Proceedings of IUI'07*, January 28–31, 2007, Honolulu, Hawaii, USA.

[6] L. Gong, and J. Lai, "To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal responses," *International Journal of Speech Technology*, vol. 6, no. 2, pp. 123–131, 2003.